HITSZ-ICRC at NTCIR-11 Temporalia Task
Yongshuai Hou, Cong Tan, Jun Xu, Youcheng Pan, Qingcai Chen and Xiaolong Wang
Key Laboratory of Network Oriented Intelligent Computation
Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China
houyongshuai@hitzs.edu.cn

Introduction

This paper presents methods HITSZ-ICRC group used in the NTCIR-11 Temporalia challenge at NTCIR-11.

- Rule based method and multi-classifier voting method were used for TQIC subtask
- Relevant score weight sum and learning to rank method were used for TIR subtask

Temporal Query Intent Classification

Temporal query intent classification method:

- Rule based
  - Use Rule set to classify queries that contain time word or time-sensitive word, like "Movies 2012", "long term weather forecast".
- Machine learning
  - Train classifier for temporal query using machine learning algorithms, including logistic regression, SMO, HNB, etc.
- Multi-Classifier voting
  - Vote results from the top N best classifiers trained
- Result merging
  - Merge results gotten from rule based method and multi-classifier voting method

Feature List used:
- N-gram term of query
- POS n-gram
- Named entity
- Normalized date
- Date distance
- Time-sensitive word

N-gram term of SE result

Temporal Information Retrieval

Temporal Information Retrieval method:

- Candidate relevant document retrieving
  - Index document set and retrieve search topic using BM25 language model
- Temporal relevant
  - Judge temporal relevant between temporal search subtopic and relevant document base the date distance of search date and time expression in the document

- Temporal search subtopic recognition
  - Use the method in subtask TQIC to recognize temporal search subtopic class
- Relevant document re-ranking
  - Re-rank relevant document list base the temporal relevant and content relevant between search subtopic and candidate document

Two methods were tried:

- Relevant score weight sum
  \[ R = \alpha R_i + (1 - \alpha) R_f \]

Learning to rank

Use date distance as rank feature for learning to rank method

Feature List for learning to rank:
- Similarity between search topic and document title
- Similarity between search topic and document content
- Similarity between search subtopic and document title
- Similarity between search subtopic and document content
- BM25 relevant score between search topic and document
- Temporal relevant score

Evaluation

Data set

- "Living knowledge news and blogs annotated sub-collection" document corpus, contains 3.8M documents from blogs and news sources.
- 300 formal run queries for TQIC testing
- 50 formal run search topics each with 4 subtopics for TIR testing

Evaluation results of subtask TQIC

Table 1: Coefficient value for relevant score weight sum method

<table>
<thead>
<tr>
<th>Subtopic Class</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>past</td>
<td>0.85</td>
</tr>
<tr>
<td>recency</td>
<td>0.73</td>
</tr>
<tr>
<td>future</td>
<td>0.76</td>
</tr>
<tr>
<td>atemporal</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 2: Results evaluation of TQIC formal runs

<table>
<thead>
<tr>
<th>runID</th>
<th>Correct Number</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW</td>
<td>207</td>
<td>69%</td>
</tr>
<tr>
<td>PrW</td>
<td>203</td>
<td>67%</td>
</tr>
<tr>
<td>qRPrHNB</td>
<td>201</td>
<td>67%</td>
</tr>
</tbody>
</table>

Table 3: Precision of each class in TQIC formal runs

<table>
<thead>
<tr>
<th>runID atemporal future past recency</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW</td>
</tr>
<tr>
<td>PrW</td>
</tr>
<tr>
<td>qRPrHNB</td>
</tr>
</tbody>
</table>

Evaluation results of subtask TIR

Table 4: Results evaluation of TIR subtask runs

<table>
<thead>
<tr>
<th>runID nDGC@20</th>
<th>AP@20</th>
<th>P@20</th>
<th>nERR@20</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW</td>
<td>0.5355</td>
<td>0.4599</td>
<td>0.5092</td>
</tr>
<tr>
<td>BWCC</td>
<td>0.4768</td>
<td>0.483</td>
<td>0.6018</td>
</tr>
<tr>
<td>LTRNC2</td>
<td>0.4544</td>
<td>0.4587</td>
<td>0.5905</td>
</tr>
</tbody>
</table>

Table 5: nDGC@20 of each class in TIR formal runs

<table>
<thead>
<tr>
<th>runID atemporal future past recency</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW</td>
</tr>
<tr>
<td>BWCC</td>
</tr>
<tr>
<td>LTRNC2</td>
</tr>
</tbody>
</table>

Conclusion

Merging results of rule based method and multi-classifier voting is effective for TQIC subtask.
Both relevant score weight sum method and learning to rank method are effective for TIR subtask, and learning to rank method is more effective here.